One hundred years of Current Employment Statistics—the history of CES sample design

To help mark the Monthly Labor Review's centennial, the editors invited several producers and users of BLS data to take a look back at the last 100 years. Since its inception in 1915, the Current Employment Statistics (CES) survey has undergone significant enhancements, including improvements to sample design. When the size of a population makes a census impractical, a representative subset of the population—a sample—is used to estimate characteristics of the population. It is important that the sample be representative of the population. The methodology that is used to select a representative sample from a population is perhaps the most critical step in designing a survey. This article discusses the evolution of the CES sample design.

Laura A. Kelter
cesinfo@bls.gov

Laura A. Kelter is a supervisory economist in the Office of Employment and Unemployment Statistics, U.S. Bureau of Labor Statistics.

In October 1915, what is now known as the Current Employment Statistics (CES) survey published employment and payroll data for four manufacturing industries—boots and shoes, cotton goods, cotton finishing, and hosiery and underwear—using data collected from fewer than 200 large manufacturing firms. The program gradually expanded in sample size and industry coverage over several decades, with the last major sample size expansion completed in the mid-1980s. This increase was undertaken, specifically, to provide adequate sample to expand publication detail in the services industries. As a result, and after 100 years of research, testing, and survey improvements, the CES program now surveys approximately 146,000 businesses and government agencies to estimate employment, hours, and earnings of workers on nonfarm payrolls. Since these businesses and agencies may have more than one physical location, they cover approximately 623,000 individual worksites.[1]

Origins and evolution of the CES sample design

In its infancy, the CES program first published estimates without a formal sample design. Soon after, CES started partnering with state workforce agencies (SWAs), beginning in 1916 with the state of New York, to conduct the CES as a federal–state cooperative program. This meant that firm data collected by individual states would be

shared with CES for the production of estimates at the national level. Other states gradually joined the cooperative program, and, by 1949, all states had agreements with CES. CES provided funding and estimation guidelines, and states collected sample data and produced state-level estimates. Before all states were covered in the cooperative program, CES collected data and made estimates for nonparticipating states.

The inception of the survey predated the development of probability sample design as an internationally recognized standard for sample-based surveys. It is not entirely clear how sample size or selection were determined in the early years, although later documentation referred to the early CES as a "collection of cut-off samples" by industry. Cut-off samples deliberately exclude a portion of the population that is being measured; CES empirically derived these cut-offs over time using research on what sizes and types of samples yielded satisfactory results when compared against administrative benchmark sources.[2]

Although some of the historical details are not available, it is clear that for the first few decades of the survey there was no formal sample design. What did exist was a strategy to sample enough units to achieve a high percentage of universe employment coverage. The early decades of the CES program focused mainly on producing employment and payroll data for manufacturing industries; hence, the main sample-coverage strategy was to sample as many of the largest manufacturing firms as possible. This strategy apparently worked well enough at the time, because manufacturing industries were mostly dominated by large firms. However, once industries in the services sector began to represent a larger portion of national-level employment, there was a demand for published data in these industries as well. Many of these service-sector industries were dominated by smaller firms, and it was not as straightforward to achieve a representative sample by including just the largest firms. This issue likely helped launch the effort by CES researchers to develop new sampling guidelines in the 1940s.

The 1940s sampling memorandum series

In the late 1940s, BLS researchers began a series of tests with a goal of developing a more rigorous approach to sampling. These tests and results from testing several sample selection and estimating procedures were documented in a series of seven sampling memorandums. An important caveat was that most of the research encompassed only five manufacturing industries: cotton goods, meat packing, electrical equipment, shipbuilding, and newspapers and periodicals.

These are the major conclusions from the sampling memorandums:

- *Sampling memorandum no. 1, May 1947.* The use of an unrestricted random sample[3] with a "blow-up" estimator[4] worked poorly for estimating industry employment. It required a sample of 90 percent or more of the universe in order to achieve acceptable results, where acceptable results were defined as a 2-percent error at a 95-percent confidence interval.
- *Sampling memorandum no. 2, November 1947; and memorandum no. 3, December 1947.* A random sample with a ratio estimator (or an estimator that uses the relationship between two variables) was determined to be preferable to the blow-up estimator for most industries because it required a smaller sample size to achieve the same level of sampling error.[5] This held if the coefficient of variation (or the ratio of the standard deviation to the mean) was greater than 0.5, which was the case for each of the tested industries with the exception of the shipbuilding industry.

- *Sampling memorandum no. 4, April 1948.* Stratification, or the dividing of the population into subpopulations by size of establishment, greatly reduced variance (how far the data are spread out) and therefore the required sample size. The tests reported in this memo divided the sample into two size strata: large and small. The strata boundaries were empirically determined by trying different boundaries to see which resulted in the smallest possible sample for a given level of variance. The large size stratum was a take-all sample, where all units are included in the sample, and the small size stratum was an unrestricted random sample with a blow-up estimator. Results from the two strata were added together to arrive at the industry-level employment estimate.
- *Sampling memorandum no. 5, July 1948.* An estimate developed from two size strata with a ratio estimator generally outperformed an estimate made using a blow-up estimator. In this iteration, one nonmanufacturing industry, miscellaneous retail trade, was included in the testing.
- *Sampling memorandum no. 6, November 1948.* Tests with sampling proportional to size, in which the selection probability of each unit is set to be proportional to its size, proved a much less efficient allocation of the sample than using a take-all approach for the large size class. Tests also were conducted using an optimum allocation method that minimized sampling error for the sample sizes being considered. This basically resulted in the same conclusions as sampling memo 5, indicating a take-all approach for the larger size class worked best to minimize variance.
- *Sampling memorandum no. 7, March 1949.* This memo reported on the results of additional tests of drawing samples and making estimates for the original five manufacturing industries. It concluded that use of ratio estimators for samples allocated proportionately across two strata was not consistently better than use of a ratio estimator based on an unrestricted random sample.

A followup internal memorandum in 1950 presented results of a series of additional tests. These tests used hypothetical data with a universe of five establishments and a sample of three establishments to illustrate variance and bias (or error from the sampling, collection, or estimation processes) of different sample designs and estimators. The same designs and estimators that were tested in memorandums 1–7 were used, and the results illustrated the same conclusions that had been reached using the empirical data.

While this series of memos set the stage for the CES program's use of a certainty size class for large establishments and the use of a ratio estimator, internal memos and American Statistical Association (ASA) papers recognized that ratio estimators contained bias and that there was an overall downward bias in CES employment estimates.

1964 sample redesign

A June 1963 internal paper titled "Analysis of BLS-790 sample with a proposal for a new sample design" was the catalyst for the first formal updates to sampling methods since the principles that were set out in the 1940s sampling memo series. The analysis was undertaken because there were newly available employment tabulations for 1959 by establishment size from the benchmark source for CES—the BLS ES-202 data.[6] This marked the first comprehensive set of ES-202 data by size class. The analysis reinforced major conclusions from the sampling memo series. The most notable conclusion was that most industries were dominated by either large or small firms and, therefore, would have a skewed distribution by firm size. The analysis also reaffirmed that a form of optimal allocation—two size strata and a ratio estimator—would work well for most industry series. These conclusions

were based on empirical observation and experimentation using the ES-202 data. The research tested a "full stratification" by 4-digit Standard Industrial Classification (SIC) and eight size strata, but the resulting 32-cell framework was considered impractical, as it would be too difficult to adequately populate all the estimating cells. As a result, cells were collapsed to yield a more practical design.

The 1963 paper recommended retaining all current sample members because of the cost, risk, and difficulty of attempting to solicit an entirely new sample. Because the new sample design would have retained the current sample, rather than using a randomly selected sample, it would not qualify as a probability sample. This meant that standard sampling variances could not be calculated, nor could standard inferences be made about the reliability of the resulting estimates. Additional recommendations were also developed on the basis of what was considered practical within existing resources. For example, if an industry was dominated by large firms (where 90 percent or more of its employment was in establishments with 250 or more employees), there was no sampling of firms with 0–19 employees; the overall sample coverage goal was determined to be 50 percent of universe employment. For industries dominated by small firms (where 75 percent or more of employment was in establishments with 10 or fewer employees), empirical experiments suggested that a 5-percent employment universe coverage was sufficient.

The paper concluded that this new design would be an improvement over the old cut-off sample design, since it laid out a balanced representation across all size classes while giving more weight to larger establishments, and because every unit would have a known probability of selection. The new design was considered a step toward a probability sample, but not an actual probability sample. The paper also concluded that the existing sample coverage in manufacturing appeared to be adequate for reliable estimates and approximated an optimum allocation design, but that many of the nonmanufacturing industries had inadequate samples.

CES began implementing the proposed new design from the 1963 paper in July 1964 with the issuance of a memorandum to regional commissioners and contract state agencies titled "BLS Employment Instructions No. B-29 New Sampling Procedures." This memo contained a set of sampling tables by SIC (4-digit SIC code in manufacturing and 3-digit SIC code in nonmanufacturing) and six size classes. The tables contained ratios that each state was to apply against ES-202 universe listings to determine sample sizes for each industry/size stratum. The sampling ratios were based on an assessment of sample needed for reliable national estimates, but did not specify requirements for state or area estimates.

The memo described this design as a "close approximation" to a form of optimum allocation known as "probability proportionate to size" (or to the average size of establishment, in the case of CES). The design distributed a predetermined number of establishments in the total sample among sampling cells on the basis of the ratio of employment in each cell to total industry employment. Within each cell, establishments were to be randomly selected (with the major exception that existing sample units should be retained). In practice, this meant that if a cell had, for example, 20 percent of universe employment at the industry level, then it would receive 20 percent of the sample units for that industry. For most industries, the take-all stratum was set at 250 employees at the establishment level. If 90 percent or more of an industry's employment was in the two largest size classes, the rule was to take all from those cells and sample no others. If less than 10 percent of the universe employment was in establishments with fewer than 20 employees, the rule was to take no sample from establishments with fewer than 20 employees.

States were instructed to begin implementing this new design immediately, although it was noted that there were insufficient resources to fully implement the design. The existing CES sample totaled 107,000 establishments, while the new design, if fully implemented, would yield a sample of 148,000. States also were instructed to direct sample solicitation efforts to areas of greatest need. It was also noted that thereafter sample adequacy should be reviewed after each annual benchmark and solicitation efforts directed toward those industries with the greatest sample deficiencies.

The 1964 sample redesign was an improvement to the previous CES sampling methods, but the redesign also contributed to the subsequent, decades-long criticism that the CES sample was potentially unrepresentative of the universe because it was not a true probability-based design. Moreover, critics noted that it was skewed toward including large firms at the expense of small firms.

1970s and 1980s quota sampling design and procedures

Sample initiation and most sample collection continued to be the responsibility of each of the 53 SWAs in the federal–state cooperative program through the 1980s.[7] Sample requirements developed by CES, in the form of ratios by industry and size (for example, select 1 out of every 10 establishments), were given to the SWAs. The original sampling ratios were developed for the 1964 memorandum cited above; later these ratios were commonly known as "G-1"—the appendix in the CES State Operating Manual that contained the ratio tables.

Individual SWAs were responsible for sample solicitation, maintenance, and monthly collection. Adherence to the quotas in G-1 was not regularly enforced or measured by CES. States were instructed to review and update their sample annually, after benchmarking, on the basis of a comparison of existing sample size versus the G-1 requirements and benchmark revision magnitudes. Each state decided for itself how large a sample it could support within its resources.

During the 1970s and 1980s, the quota sample method established in 1964 continued, basically unchanged. A few modifications were made on the basis of empirical experimentation, although there is no available record of this research. A minimum of 15 establishments or 50 percent universe sample coverage was set for all basic estimating cells. A second set of ratios, known as appendix G-4, was developed and used as a guideline for sample adequacy for statewide and metropolitan statistical area (MSA) estimates, though the origins of these ratios are unclear. Appendix G-4 contained two separate algorithms, one for states and MSAs with total nonfarm employment of at least 500,000 and one for those with total nonfarm employment of less than 500,000.

States were encouraged to oversample on the basis of their previous experience in order to account for nonresponse. If they generally had a success rate of 25 percent in soliciting new units, states were directed to sample 4 times as many units as the ratio tables called for. States were also directed to examine the magnitude of statewide and MSA benchmark revisions to determine which industries or areas might be most in need of sample supplementation.

Though the sample design did not change, the sample size greatly expanded during the 1980s. States were directed to solicit as many additional sample units as possible so that CES could expand its published series in the service-providing industries, as recommended by the 1979 Levitan Commission.[8] The sample size grew from approximately 160,000 establishments in 1975 to 425,000 establishments by 1989. The total sample size has remained nearly unchanged since then; however, the design was changed to a probability-based sample during

the CES redesign that began in 1995. To create a true probability sample, the redesign resulted in the cancellation of thousands of existing sample members and solicitation of thousands of new ones.

Motivation for change to probability sampling

Quota sampling is at risk for substantial bias; in contrast, the random selection technique inherent in probability sampling provides a measure of protection from bias. A random sample is more likely to be representative of the universe because a random selection balances out both the known and unknown characteristics of a population.

Probability sampling had been the standard for sample surveys for more than 50 years before it was adopted by CES. CES had not previously converted mainly because of the time, the expense, and the risk to ongoing operations required to change a massive, ongoing, monthly survey. Many people, both inside and outside of BLS, believed the existing sample was working well, since it generally experienced small to moderate-size benchmark revisions at the total nonfarm level for the national and statewide estimates and retained a high universe coverage (over one-third of total nonfarm employment).

A BLS-commissioned study of the CES program by a special panel of the American Statistical Association issued a report in 1994 that had, as a primary recommendation, the adoption of a probability sample design. BLS had commissioned this panel after an unusually large benchmark revision for March 1991 that caused extreme concern among major data users and within BLS. Although CES sample design deficiencies were not found to be the cause for this large benchmark revision, this report helped provide the incentive for CES to begin work on a new sample design.[9]

In addition, internal CES research from the mid-1990s demonstrated the bias inherent in the existing quota sample by comparing the age distribution of firms in the CES sample against the age distribution of establishments in the universe for 10 large states. On average, the CES sample units were approximately 10 years older than the universe as a whole. This research also simulated estimates made for the continuing sample units' portion of the universe, using the existing CES sample.[10] The results showed very large benchmark revisions, indicating that the existing CES sample was not performing well and that techniques such as bias adjustment (at the national level) and manual data interventions (at the state and MSA levels) were being used to override sample results to a large degree. These techniques were used in an attempt to compensate for the nonstandard sample design, as well as to account for business births. Historically, these techniques had varying degrees of success, as demonstrated by the varying sizes of benchmark revisions. This research helped convince the SWAs, as well as BLS top management, that quota sampling was no longer viable.

On the basis of its internal research results and the ASA report, CES undertook an extensive research effort to develop the most optimal sample design that would be both affordable and technically defensible. CES included outside experts in survey design and estimation in its research. These experts included senior statisticians from Westat (a private research firm), the National Opinion Research Center at the University of Chicago, and the University of Michigan Survey Research Center.

As part of the research, simulations were completed for all states and distributed to SWAs to familiarize them with expected results from a new sample design. Additional work was completed on improving sample solicitation and data collection techniques needed to sustain an adequate response rate. Extensive research was also completed

on estimation algorithms and on sampling new business births in real time versus modeling for their contributions to total employment.

Current probability sample design

Development of the CES probability sample design began in 1995, was completed in 1997, and was production tested between 1998 and 2002. The implementation of the new sample design into production was phased in between June 2000 and June 2003, following successful production tests. Each June, one or more major industry divisions were converted to the new sample at the time of the annual benchmark revision. Wholesale trade was converted in 2000; mining, construction, and manufacturing in 2001; transportation and public utilities, retail trade, and finance, insurance, and real estate in 2002; and finally, services in 2003. At the time, the industry classification system was SIC—the precursor to the North American Industry Classification System (NAICS).

Since CES was already obtaining near universe coverage for federal and state government, a probability sample for government was not developed. Local government is a possible candidate for future conversion to a probability sample, although coverage in that sector is also very high. The current local government sample is a legacy sample from the old quota sample design, when individual states were responsible for sample selection and collection.

Sample size and allocation by state

As a starting point, the first CES probability sample (drawn in 1998) set the new probability-based sample size to be equal to the existing quota sample size for each state. This was done to obtain the cooperation of SWAs in the transition to a new design and to preserve the approximate overall sample coverage. BLS also determined that this was the total sample size that the program resources could support. This effectively resulted in a state-based design.

During the sample redesign research phase, both national and MSA-based designs were also considered; however, the state-based design was found to be the most efficient in terms of balancing several program goals. A national design would require very little sample from the smallest states, leaving them with unacceptably large variances. By contrast, an MSA-based design would be very inefficient for producing national estimates and would increase sampling error to an unacceptable degree. The top two priorities for the CES program are reliable all-employee estimates at the national and statewide levels, followed by MSA-level estimates. A state-based design was determined to be the best option to satisfy these competing priorities.

State sample reallocations

Approximately 5 years after the initial CES redesign was completed, BLS reevaluated the sample size and its distribution across states for the first time and completed a relatively minor reallocation across states. The total sample size was slightly increased at this time. The sample review examined several possible methods for determining the sample distribution across states. One method involved equalizing relative standard errors (RSE) across states. This approach was deemed unacceptable because it resulted in a large increase in variance for national and large states' estimates. CES also examined an unbounded optimum allocation to produce a national total nonfarm employment estimate with the smallest sampling error. This approach was unacceptable because it substantially increased variance for smaller states. The final allocation selected was a bounded, national optimum allocation, which specified upper limits on states' RSEs and limits on how much sample any individual state could

lose. The maximum loss was set at 15 percent, and the resultant national RSE of .075 percent was considered acceptable. The practical effect of the reallocation was primarily to reallocate sample toward employment-growth states and away from states with declining employment. The final allocation is close to a proportional allocation, where each state's allocation is approximately equal to its proportion of national total nonfarm employment (with minimum sample limits set for the smallest states).

Another minor reallocation was completed in 2011 on the basis of similar principles. CES plans to continue to review and reallocate the sample across states approximately every 5 years.

Sampling methods

The CES sample is a stratified simple random sample of worksites clustered by unemployment insurance (UI) account number. It is a state-based design where each state is assigned a fixed sample size, and then an optimum allocation procedure is used to allocate the total sample across CES supersectors[11] and size classes to achieve the maximum level of reliability for the total nonfarm employment estimate for the state.

In general, an optimum allocation distributes a fixed number of sample units across a predetermined set of strata to minimize variance (also known as sampling error) on the primary estimate of interest. For CES, the primary estimate of interest is the total nonfarm monthly employment change. The optimum allocation assigns more sample in cells for which data are less costly to collect, cells that have more units, and cells that have a larger variance. The sampling rate for each cell is a result of the optimum allocation.

Timing and procedures for sample updates

The entire sample is redrawn annually in the fall of each year from the most recent first-quarter Longitudinal Data Base (LDB), which is a longitudinally linked database file from the CES administrative benchmark source—the BLS Quarterly Census of Employment and Wages. This update incorporates new business births and updates control information on industry and geographic classifications for existing firms. The annual update also removes business deaths from the frame. In addition to the annual full sample redraw, a sample of new business births is drawn midway through each year. The sampling frame for births includes only those firms born since the first-quarter sample draw.

These updates keep the sample as current as possible with respect to business births and deaths. Until 2014, about 1 full year separated the annual sample draw and the introduction of the sample into live monthly estimates; this time was needed to complete sample refinement, enrollment, collection, and validation of initial reported data values. As an example, monthly estimates for January through December 2013 were made using sample selected from the 2011 LDB file. Beginning in August 2014, the annual updates were introduced under a quarterly sample implementation pattern by supersector that allows the sample to be included in estimation once the enrollment for the supersectors is completed. There is still a lag in the sample availability, relative to the current month's estimate; however, quarterly implementation decreases that lag compared with the previous annual sample implementation.

Because of the lag in picking up business births and deaths in the live monthly production sample, CES uses a model-based technique to account for the contribution of recent business births and deaths in the monthly estimates. For more information on the CES birth/death model, see https://www.bls.gov/web/empsit/cesbd.htm.

Because of the cost and workload associated with enrolling new sample units, all units remain in the sample for a minimum of 2 years. To ensure that all units meet this minimum requirement, CES established a "swap" procedure. This procedure allows units to be swapped into the sample that were newly selected during the previous year but not reselected as part of newly drawn sample for the current year. The procedure removes a unit within the same selection cell and places the newly selected sample from the previous year back into the sample. Approximately two-thirds of the drawn sample overlaps from one year to the next, much of it in the large size classes.

As another practical consideration, most firms are kept in the sample for no more than 4 years, except for certainty size class reporters. This is done to minimize respondent burden and help elicit cooperation when a firm is first solicited for the CES survey.

The CES sample is drawn at the UI account level within each state, but CES collects data at the establishment level within each UI account whenever possible. The establishment-level breakouts are primarily an aid in making more accurate MSA estimates. Some UI accounts contain establishments with multiple industry (NAICS) codes, and thus the establishment-level breakouts contribute to more accurate industry estimates as well at national, statewide, and MSA levels. For firms that do not provide worksite-level breakouts, a disaggregation method known as proration is used to distribute the UI account-level data to the worksite level.

Sample coverage by industry

The latest benchmark employment levels and the approximate proportion of universe employment coverage at the national total nonfarm level and at NAICS supersector levels can be found on https://www.bls.gov/web/empsit/cestn.htm#Tb1. The sample distribution by industry reflects the goal of minimizing the sampling error on the total nonfarm employment estimate, while also providing for reliable estimates by industry. Sample coverage rates vary by industry as a result of building a design to meet these goals.

Conclusion

The modern-day CES program evolved from one that began 100 years ago, when employment estimates were published for four manufacturing industries before there was a formal sample design. Through research, testing, and cooperation with state partners, the program has undergone significant improvements in sample design that have not only improved reliability of the estimates, but also helped make CES data one of the world's most-watched principal federal economic indicators. Those improvements include increased coverage to better represent the economy, enhanced representation of both small and large firms, and reduced costs and respondent burden. Today, the CES sample is a stratified, simple random sample of UI accounts (and all of their associated establishments), a design that reflects contemporary standards for sample-based surveys.

SUGGESTED CITATION

Laura A. Kelter, "One hundred years of Current Employment Statistics—the history of CES sample design," *Monthly Labor Review,* U.S. Bureau of Labor Statistics, August 2016, https://doi.org/10.21916/mlr.2016.35.

NOTES

[1] The Current Employment Statistics (CES) program provides detailed industry data on employment, hours, and earnings of workers on nonfarm payrolls. For more information on the program's concepts and methodology, see "Technical notes to establishment survey

data," https://www.bls.gov/web/empsit/cestn.htm. To access CES data, see "Current Employment Statistics—CES (national)," https://www.bls.gov/ces.

2 Benchmarking is a process of aligning sample estimates with population values, usually from an administrative data source.

3 An unrestricted sample is a sample in which units can be selected with equal probability, and because sample units can be selected more than once, they can be replaced.

4 Despite research efforts, the CES program was unable to find the specifics of the "blow-up" estimator tested in sampling memorandum no. 1. Consensus among program experts is that the method was a form of a Horvitz-Thompson estimator. See D.G. Horvitz and D.J. Thompson, "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, December 1952, pp. 663–685, http://lib.stat.cmu.edu/~brian/905-2008/papers/Horvitz-Thompson-1952-jasa.pdf.

5 Sampling error is the error resulting from observing a sample rather than the entire population.

6 The ES-202, now known as the Quarterly Census of Employment and Wages (QCEW) program in BLS, became the administrative benchmark source for the CES program in 1935. Employers are required to regularly report employment and payroll data to this program.

7 There are a total of 53 state workforce agencies, but only the 50 states and the District of Columbia contribute to national total nonfarm employment.

8 The Levitan Commission, also known as the National Commission on Employment and Unemployment Statistics, was founded to examine the government's labor force statistics and recommended that the programs be reviewed at least once each decade.

9 The large revision was actually caused by a reporting change by payroll processing firms in their development of employment counts submitted to the QCEW program for benchmarking. The ASA panel validated this conclusion by CES.

10 Continuing units are defined as all units, excluding births (new) and deaths (those going out of business).

11 The CES survey publishes estimates for aggregated NAICS sectors called "supersectors."

RELATED CONTENT

Related Articles

The *Monthly Labor Review* turns 100, *Monthly Labor Review*, July 2015.

One hundred years of Current Employment Statistics data collection, *Monthly Labor Review*, January 2016.

People who are not in the labor force: why aren't they working?, *Beyond the Numbers*, December 2015.

Related Subjects

Technological change   │   Survey methods   │   Statistical programs and methods   │   Bureau of labor statistics   │   History